The development on HYBRILIT of the Machine-learning algorithms for identification and separation of the neutron and gamma-ray signals obtained from the DEMON detector

D.V. Podgainy [1], O.I Streltsova [1], Y.A. Butenko[1] , E.M. Kozulin[1] and A.I Streltsov [2]

[1]*Joint Institute for Nuclear Research, Dubna*

[2]Institut für Physik und CINSaT, Universität Kassel, Kassel, Germany

With collaboration group of
E.M. Kozulin (Flerov Laboratory of Nuclear Reactions, JINR)

HybriLIT: heterogeneous computation cluster

The cluster contains 10 computational nodes with graphical accelerators NVIDIA Tesla K20X, K40, K80, Intel Xeon Phi 7120P, 5110 coprocessors. All computational nodes include two Intel Xeon E5-2695v2 and V3 processors each.

Computation component HybriLIT

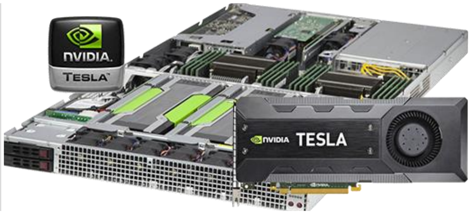
TOTAL RESOURCES

- 252CPU cores;
- 77184 CUDA cores;
- 182 MIC cores;
- ~2,5 Tb RAM;
- ~57 Tb HDD.

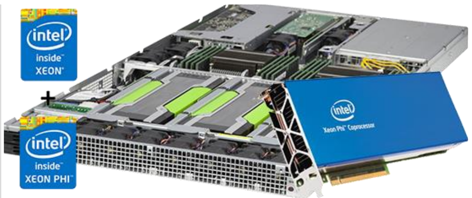
HARDWARE

SuperBlade Chassis including 10 calculation blades for run user tasks.

7 blades include specific **GPU accelerator** sets. Driven by **NVIDIA CUDA** software.



1 blade includes **2 PHI accelerators**. Driven by **Intel MPSS** software.



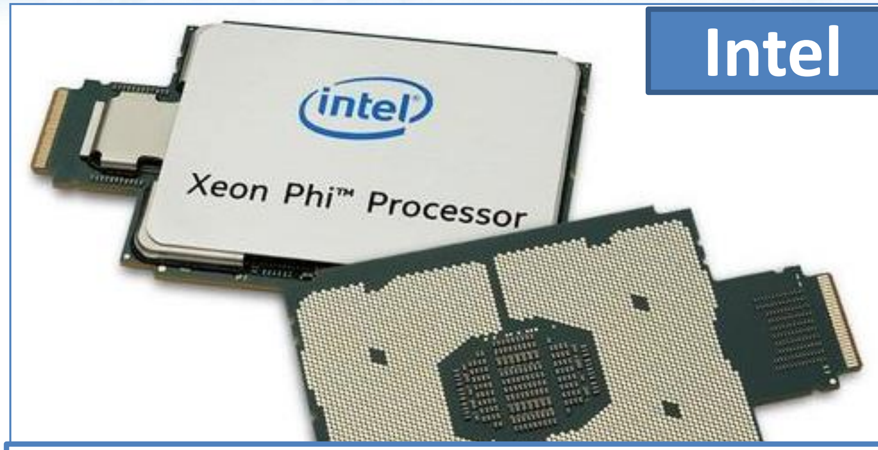
1 blade includes **1 PHI** and **1 GPU accelerators**. Mixed **NVIDIA CUDA** and **Intel MPSS** software.

1 blade includes **2 multi-core CPU processors**. Large ~7 Tb storage area

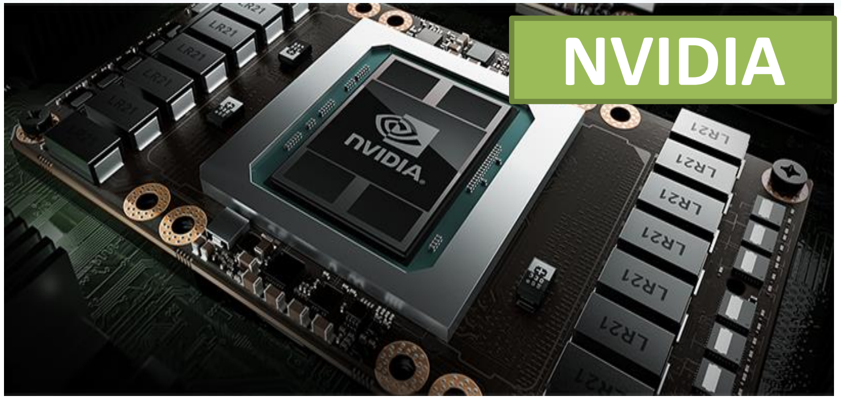
LIT

hYBRI

New computer architecture and IT-technology



Intel® Xeon Phi™ Processor 7290F
http://ark.intel.com/products/95831/Intel-Xeon-Phi-Processor-7290F-16GB-1_50-GHz-72-core



NVIDIA® Tesla® P100 GPU accelerators
http://www.nvidia.com/object/tesla-p100.html

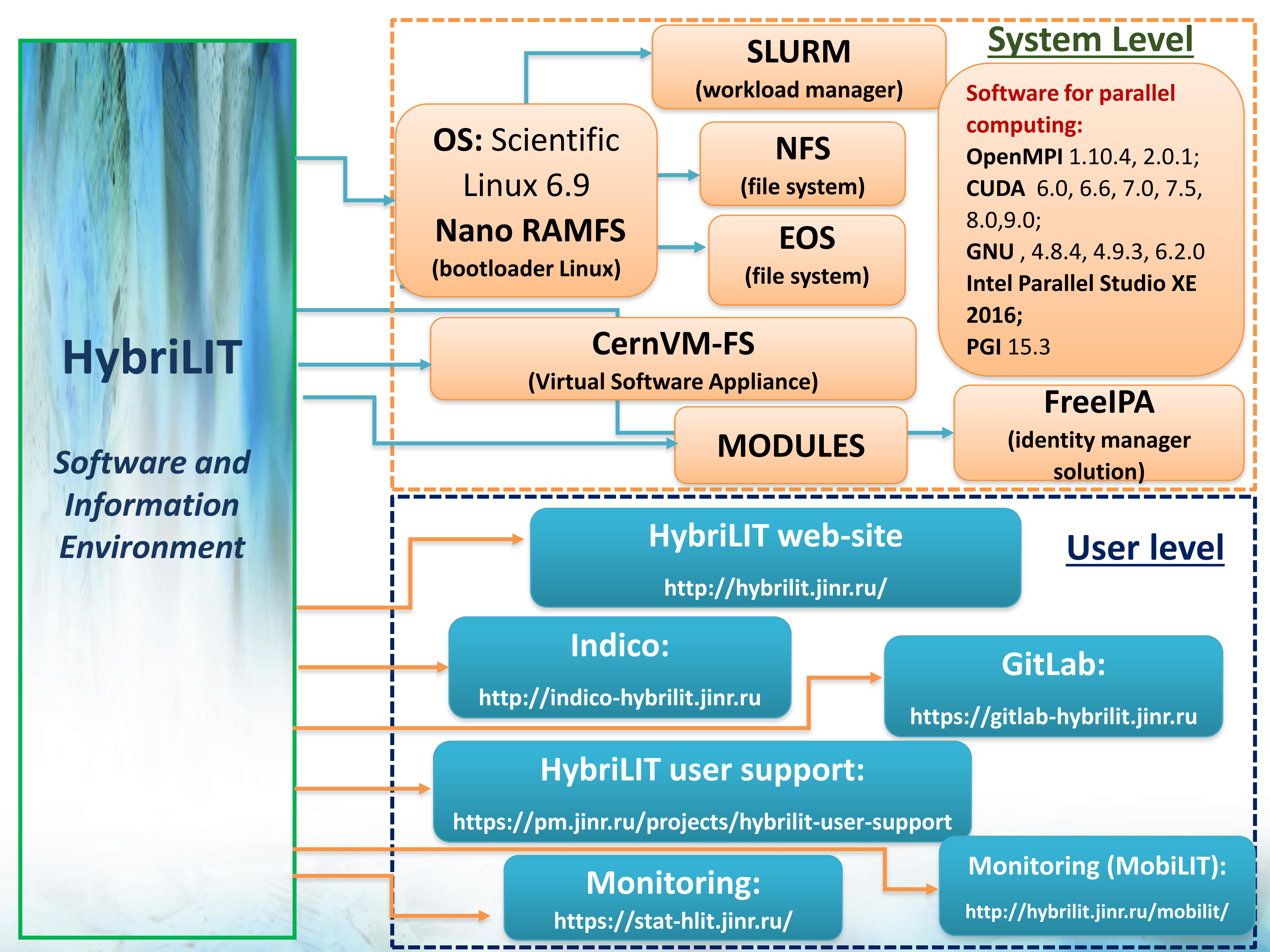
ZAO RSC Technologies
http://www.rscgroup.ru/

HybriLIT may serve

- **to create** own software for investigations demanding resource-intensive computations

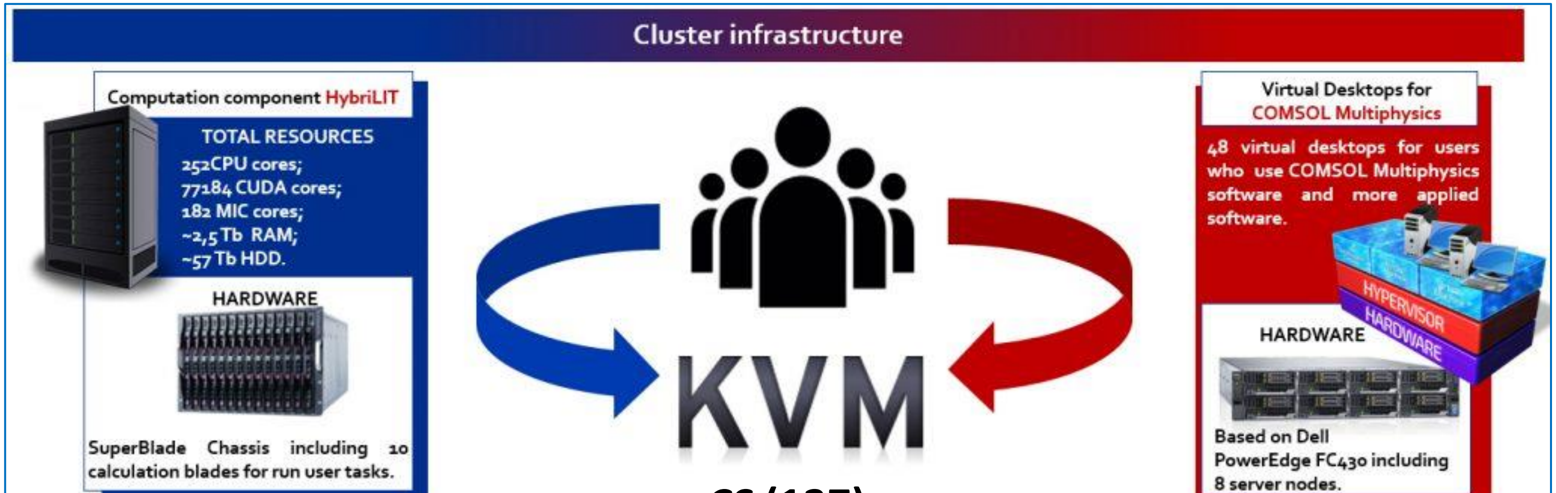
- **to use** already developed software products and applied mathematical libraries for calculations on hybrid architectures

- **to develop** parallel algorithms for experimental data processing and analysis using programming paradigms for specialized computing systems consisting of graphic accelerators and co-processors


HybriLIT
Software and Information Environment
System Level
SLURM
(workload manager)
OS: Scientific Linux 6.9
Nano RAMFS
(bootloader Linux)
NFS
(file system)
EOS
(file system)
Software for parallel computing:
OpenMPI 1.10.4, 2.0.1;
CUDA 6.0, 6.6, 7.0, 7.5, 8.0,9.0;
GNU , 4.8.4, 4.9.3, 6.2.0
Intel Parallel Studio XE 2016;
PGI 15.3
CernVM-FS
(Virtual Software Appliance)
MODULES
FreeIPA
(identity manager solution)
User level
HybriLIT web-site
http://hybrilit.jinr.ru/
Indico:
http://indico-hybrilit.jinr.ru
GitLab:
https://gitlab-hybrilit.jinr.ru
HybriLIT user support:
https://pm.jinr.ru/projects/hybrilit-user-support
Monitoring:
https://stat-hlit.jinr.ru/
Monitoring (MobiLIT):
http://hybrilit.jinr.ru/mobilit/

HybriLIT IN 2016: NEW POSSIBILITIES AND SERVICES

New possibilities for carrying out computations: a component for using COMSOL Multiphysics



CS (137)

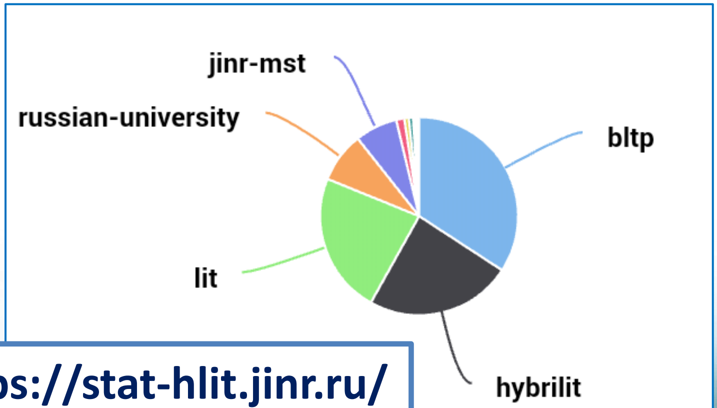
New services providing a more convenient workspace for computations: *http://hybrilit.jinr.ru*

MobiLiT – is a native Android mobile application for users of the HybriLIT HPC cluster.

Developer: Alexej I. Streltsov (Heidelberg, Germany), developed with support from HybriLIT team.

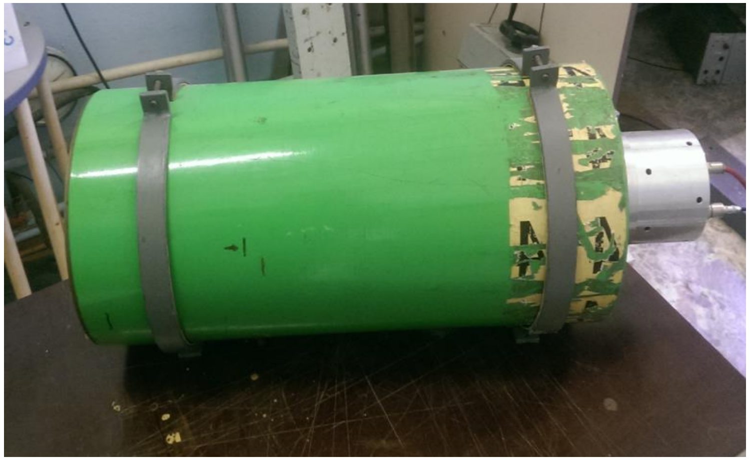
We are available on **Google Play Store** as **“MobiLIT@HPC”**.

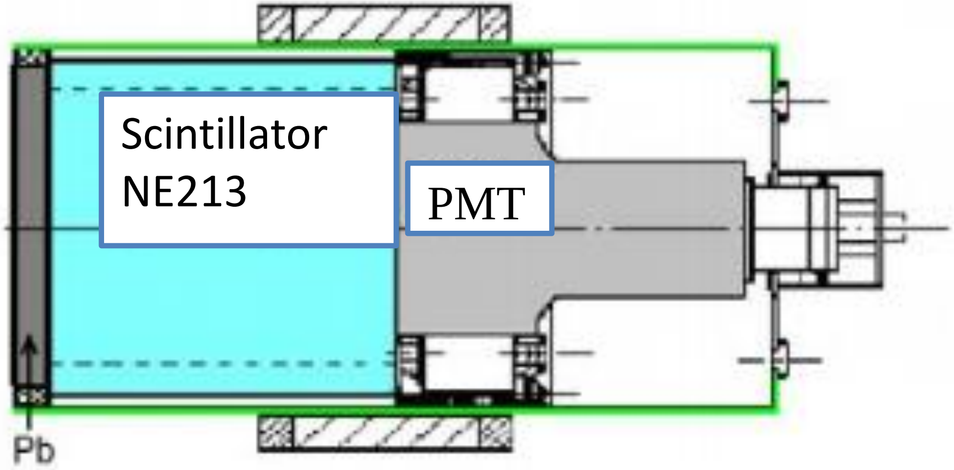
Stat-HybriLIT– is a web-service for information support of users of the HybriLIT:



https://stat-hlit.jinr.ru/

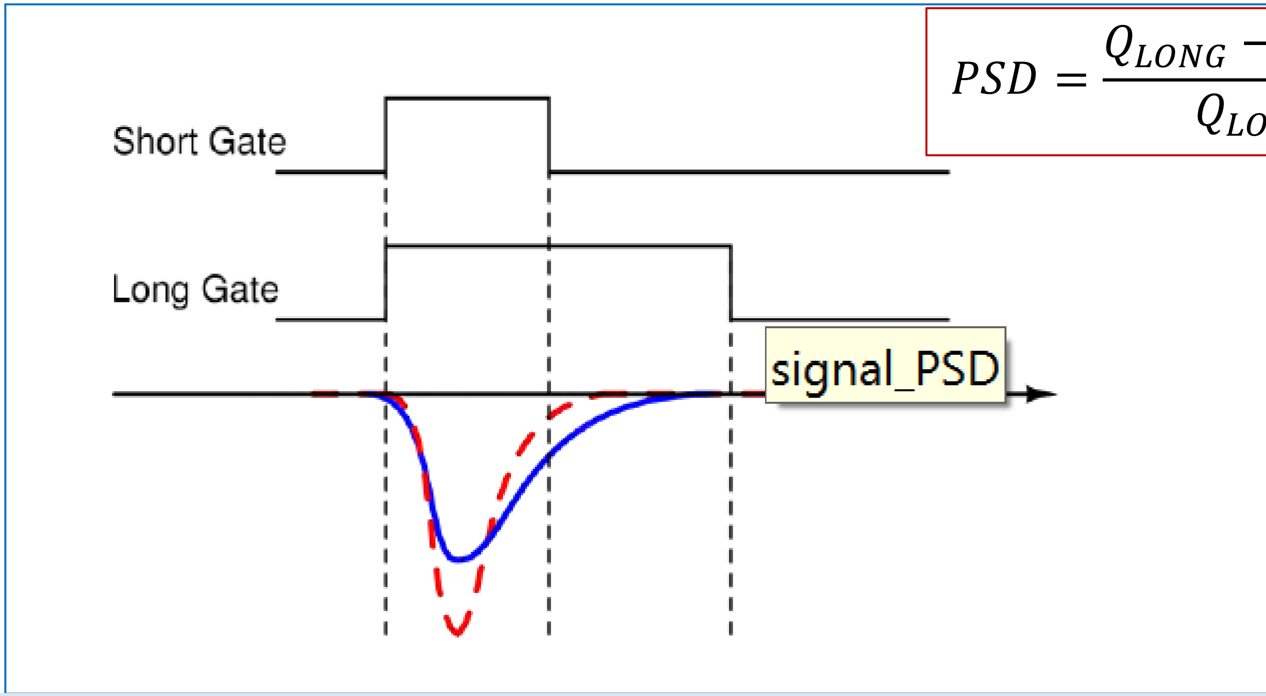
Spectrometer *DEMON* (*DE*tecteur *MO*dulaire de *N*eutrons)





The **DEMON** neutron detectors use a ***NE213*** organic scintillator, which is coupled to a XP4512B photomultiplier tube. The effective diameter of the scintillator cell is 16 cm and the length is 20 cm. The effective diameter of the photo-cathode of the photomultiplier tube is 11 cm. A thick layer of aluminum tube (2 cm thick) is added between the cell and the outside housing of a 20 cm diameter steel tube in order to minimize the cross talk between detectors.

DEMON: Method of signal discrimination

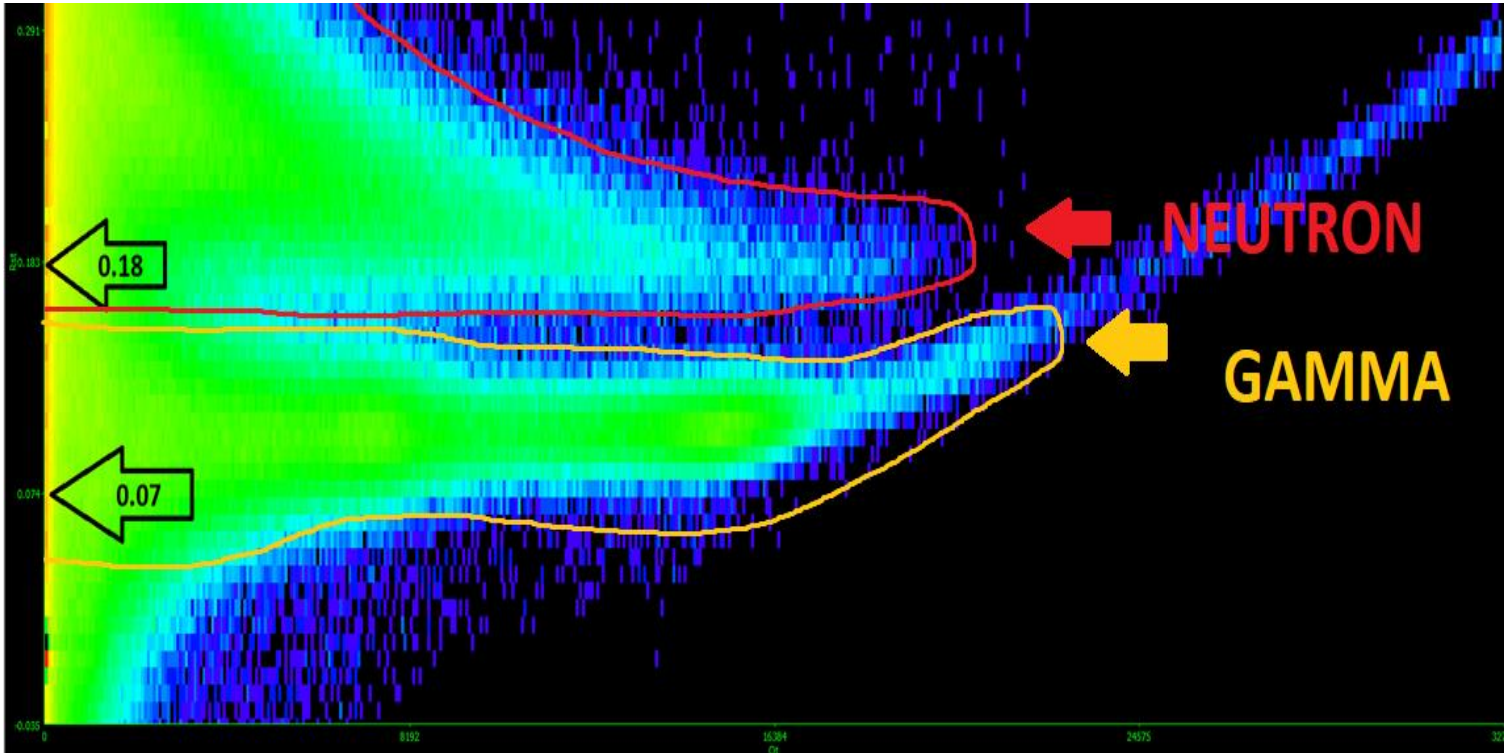


$$PSD = \frac{Q_{LONG} - Q_{SHOT}}{Q_{LONG}}$$

Fig. From **Digital Pulse Processing for Pulse Shape Discrimination (DPP-PSD) Control Software User Guide:** http://www.caen.it/csite/CaenProd.jsp?idmod=802&parent=38#

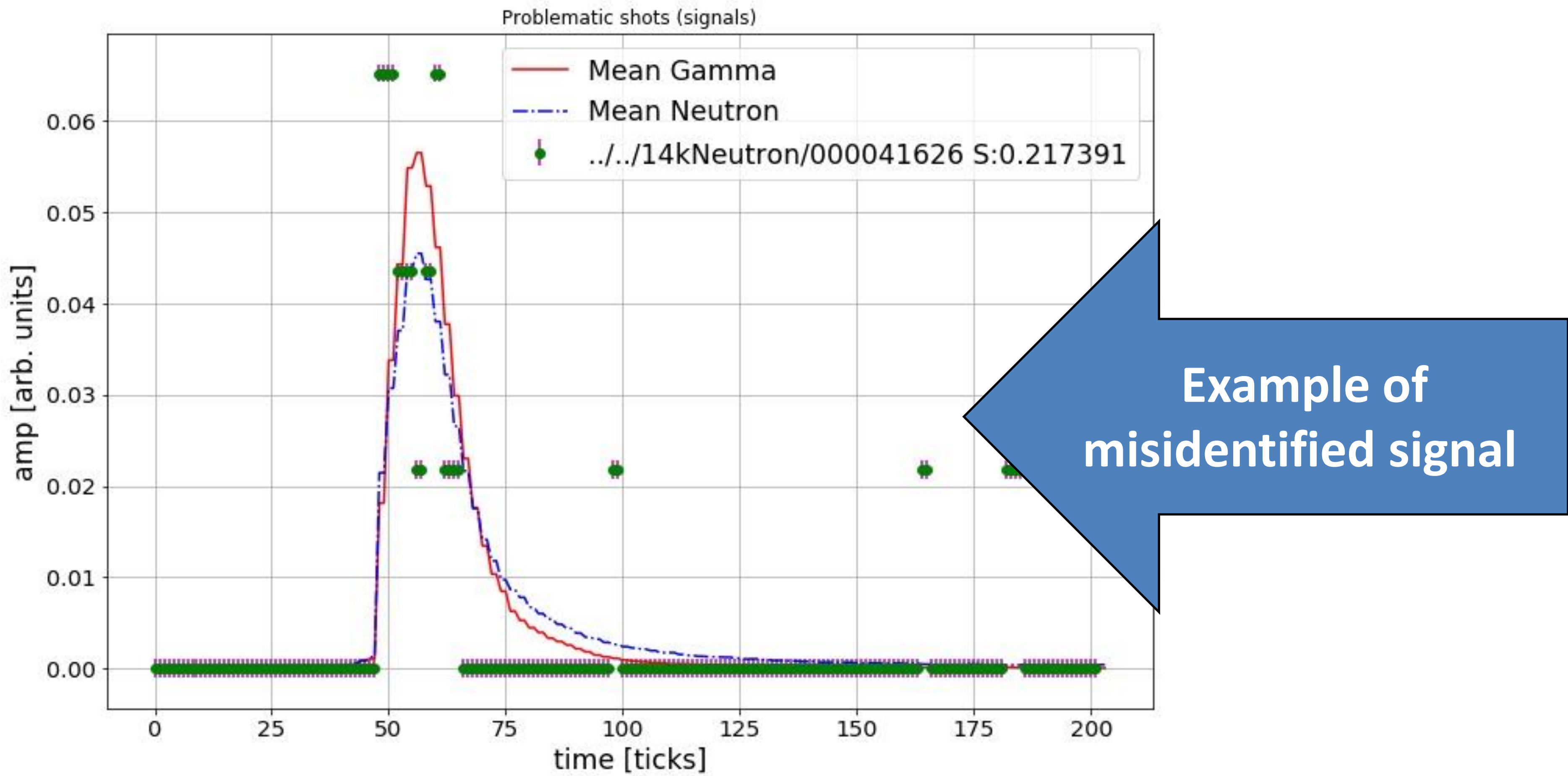
The aim of the DPP-PSD firmware is to calculate the two charges Q_{SHOT} and Q_{LONG}, performing a double gate integration of the input pulse. The ratio between the charge of the tail (slow component) and the total charge gives the PSD parameter used for the gamma-neutron discrimination.

DEMON – Problem of signal separation #1



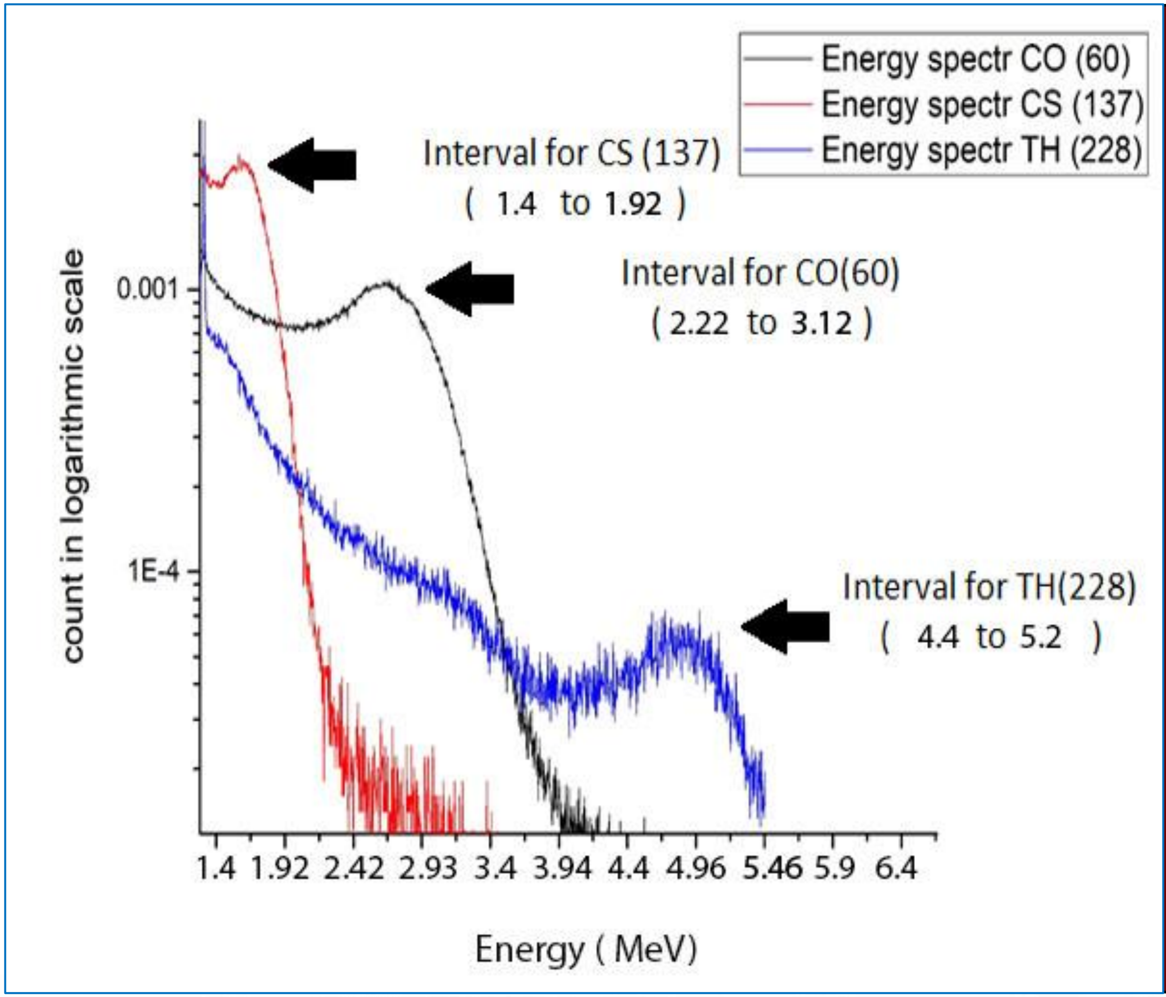
We see that the method of double gate integration cannot successfully successfully discriminate the gamma and neutron signals over the whole energy interval. However, we can use it as a pre-sort input for ML.

DEMON – Problem of signal attribution #2



We see that the double gate integration method mistakenly attributes the signal of a noise "ejection", as a neutron.

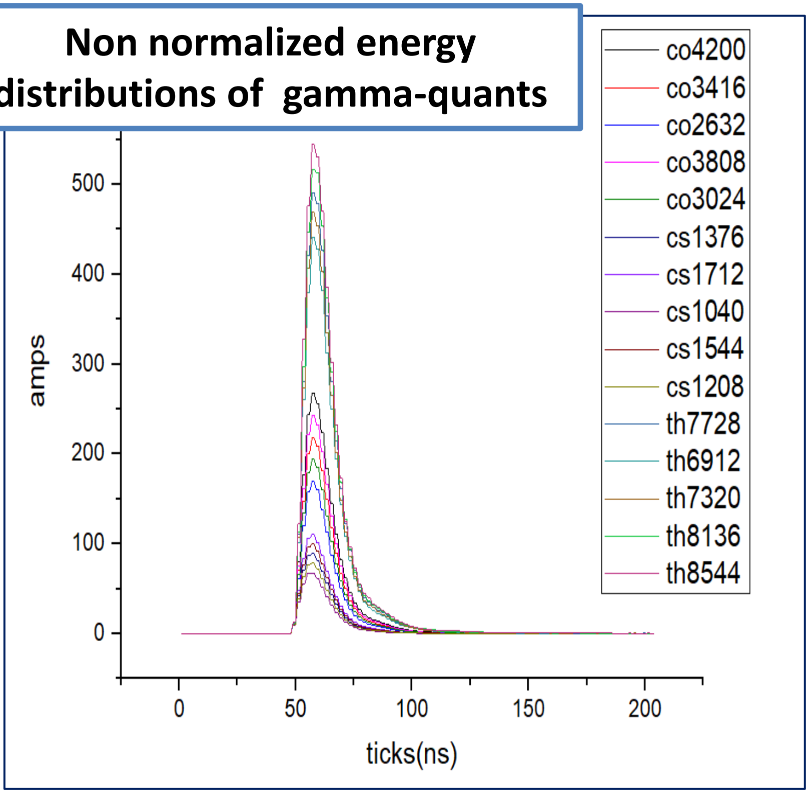
Energy distribution for gamma sources used

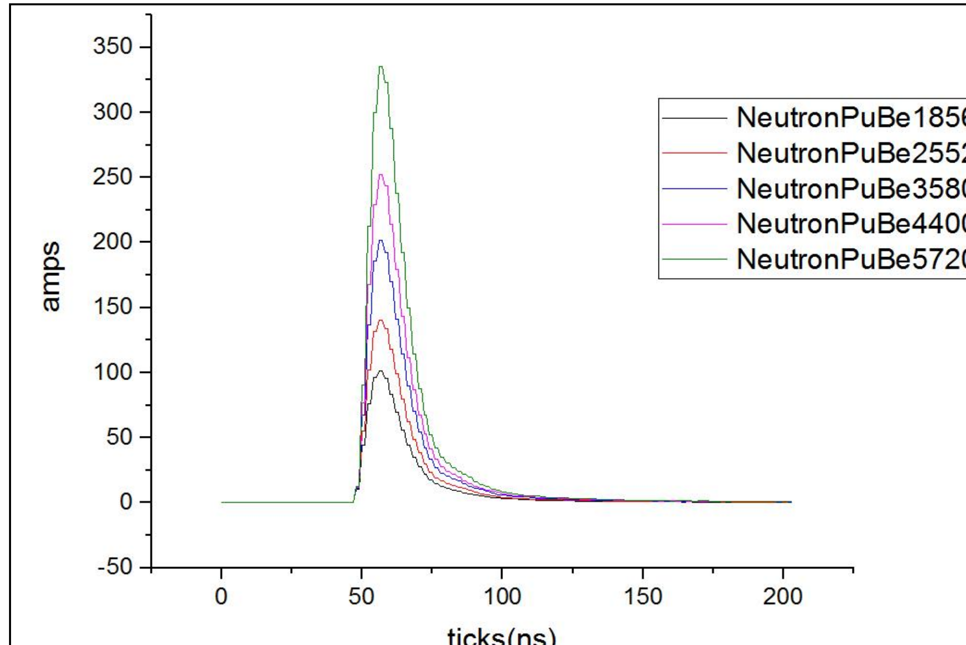


From this graph, we can see in which intervals of energies (channels) the maximum of the detected signals for three sources lay. On the basis of the presented graph we define the interval boundaries for the ***"reliable"*** gamma-quants.

Gamma and Neutron sources of radiation

Non normalized energy distributions of gamma-quants



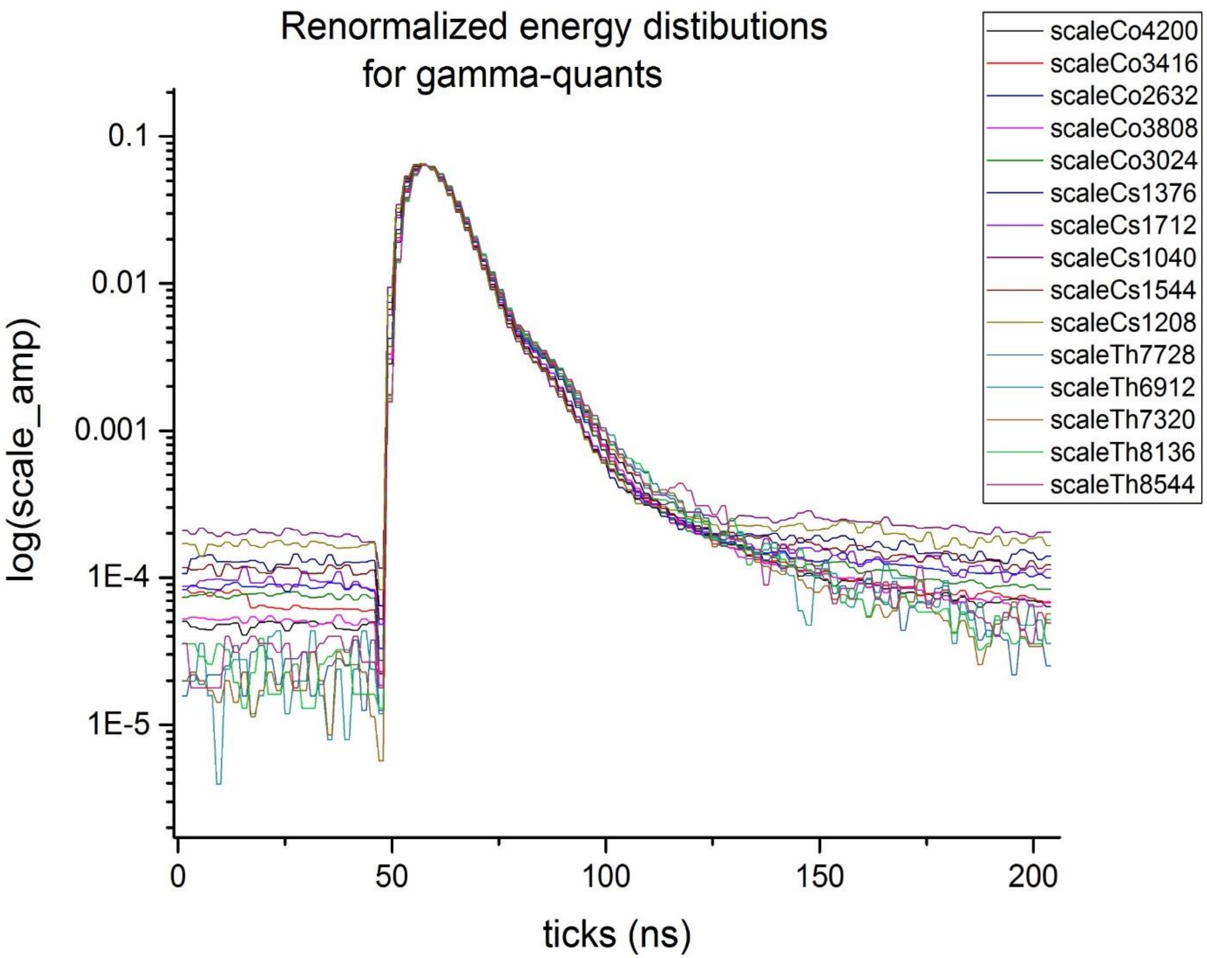


Non normalized energy distributions of neutrons from PuBe double source

***Co* (60)** – 2632, 3024, 3416, 3808, 4200,
***Cs* (137)** – 1040, 1208, 1376, 1544, 1712,
***Th* (228)** – 6912, 7320, 7728, 8136, 8544;

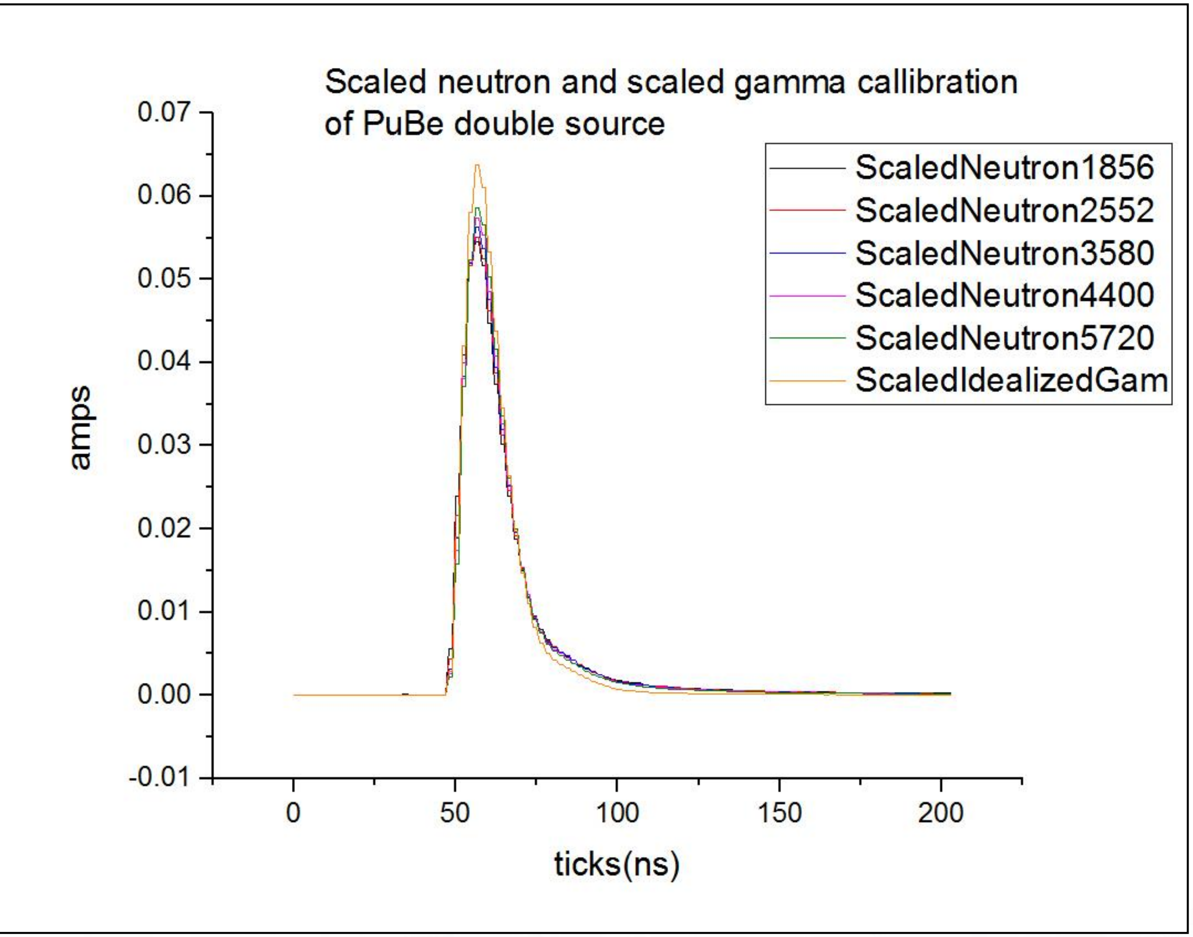
PuBe – 1856, 2552, 3580, 5720

Renormalized energy distribution for gamma-quants



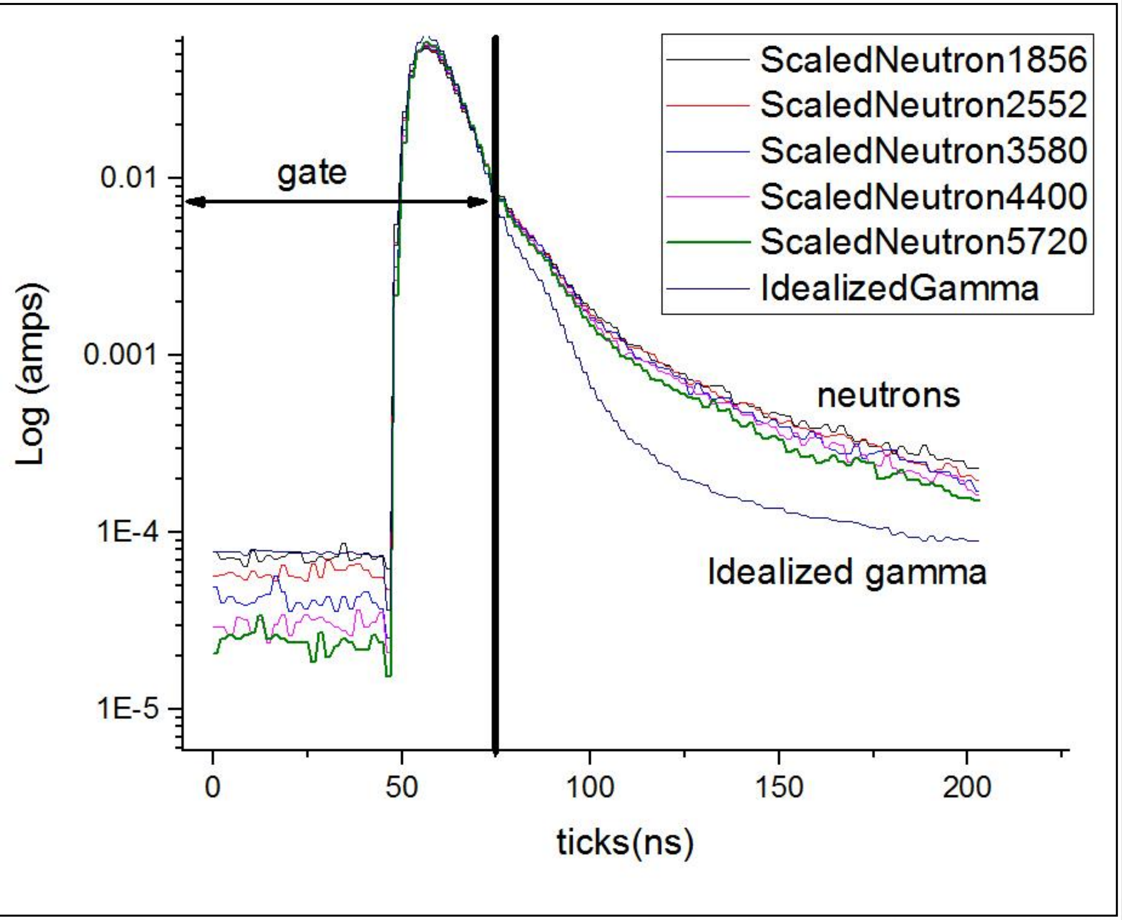
The main observation is that the shots (experimental detections) of all gamma-quants have the same shape/profiles. This fact allows us to introduce a concept of an ***idealized gamma-quant*** – by renormalizing all the gamma-quant signals by their energy. We rescale the detected profiles by the corresponding total energy (sum of all amplitudes). The above figure confirms our observations.

Renormalized energy distributions for neutrons and idealized-gammas



The graph shows that the scaled neutron signals ***differ*** from the ideal gamma profile: neutrons have a lower maximum, but broader, i.e., more pronounced tail (see next slide) .

Previous analysis (with DEMON separating method)



Renormalized neutron and gamma-shots in logarithmic scale. Gate marks the region used in previous analysis to separate gamma from neutron-signals according to the square under the tail rule:

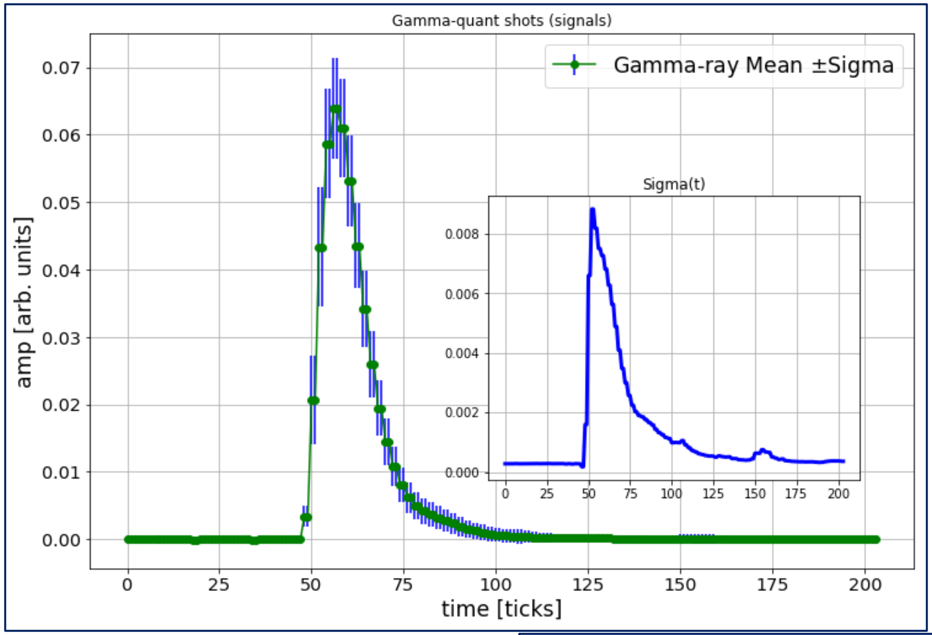
Gamma if

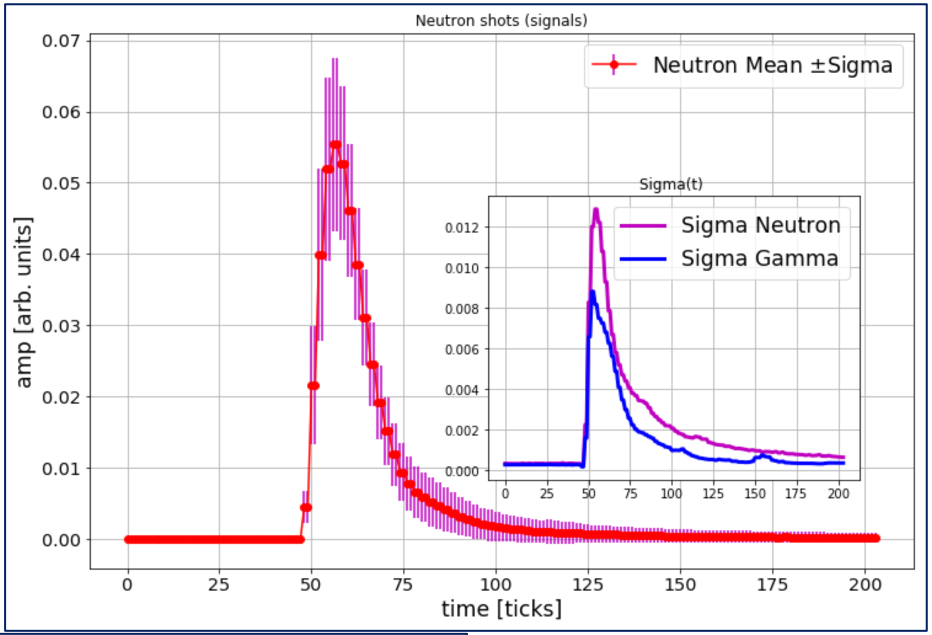
$$\frac{S_{total} - S_{gate}}{S_{total}} < 0.1$$

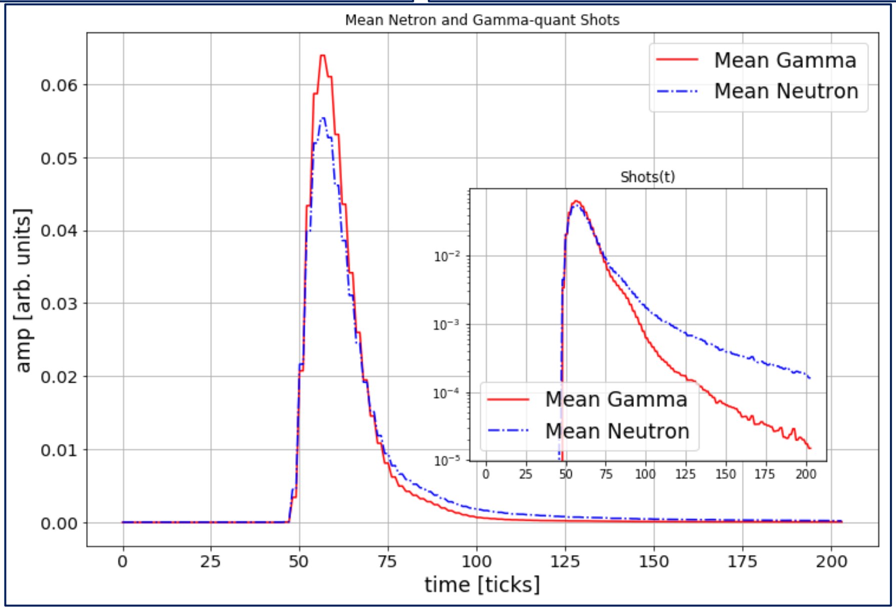
and **Neutron** if

$$\frac{S_{total} - S_{gate}}{S_{total}} > 0.2$$

Normalized profiles of gamma- and neutron-signals: Statistical analysis.







SOFTWARE for ML

Anaconda is a free open source distribution of the **Python** and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment

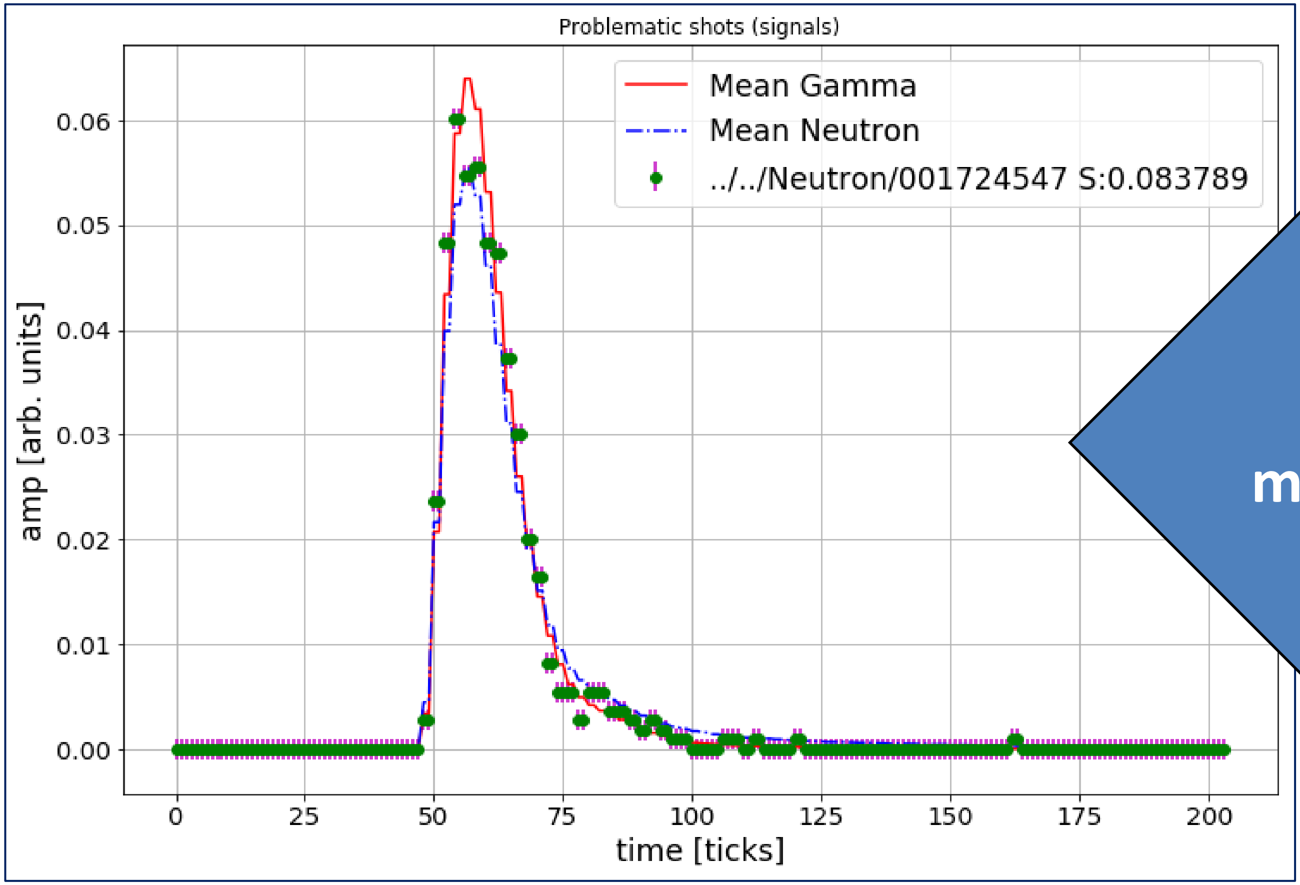
The **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Scikit Learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries **NumPy and SciPy**

Result of ML classifiers for pre-sorted dataset

Classifier	Seconds	Score
Linear-SVM	44.58	0.49341
RBF-SVM	29.56	0.79002
Naive-Bayes	0.13	0.78691
Decision-Tree	0.32	0.97336
Random-Forest	46.31	0.98991
AdaBoost (50 parameters)	2.76	0.94466
AdaBoost1 (200 parameters)	10.96	0.99922

Misidentified signals (with AdaBoost1 classifier)



Example of misidentified signal

Next step is the «purification» of the data, by excluding misattributed shots from the consideration (we use the predictions obtained within the AdaBoost1 classifier)

ML results after data-«purification»

Total number of shots excluded after “purification”:

Gamma = **80**

Neutron = **87**

Train length = **7836** signals.

~ 2.1% of signals was excluded from the dataset.

ML results after data-«purification»

Classifier	Seconds before	Seconds after	Score before	Score after
Linear-SVM	44.58	44.48	0.49341	0.49384
RBF-SVM	29.56	28.52	0.79002	0.79456
Naive-Bayes	0.13	0.12	0.78691	0.79378
Decision-Tree	0.32	0.38	0.97336	0.97202
Random-Forest	46.31	46.29	0.98991	0.99378
AdaBoost (50 parameters)	2.76	2.70	0.94466	0.94845
AdaBoost1 (200 parameters)	**10.96**	**10.72**	**0.99922**	**1.00000**

After data «purification» Adaboost1 has the best score. So, in our opinion, it is the most suitable machine learning algorithm for separation and classification of gamma- and neutron-signals

Conclusions

- We apply several machine-learning (ML) algorithms for identification and separation of the neutron and gamma-ray signals coming from the DEMON detector

- The ML-predictions have been contrasted with the results obtained within a standard method based on an integral-area scheme. In the situations where the standard method fails a properly trained ML-algorithm provides more adequate predictions and, therefore, performs much better.

Thank you for attention!

Join HybriLIT!